Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **461,** 3432–3442 (2016) doi:10.1093/mnras/stw1503
Advance Access publication 2016 June 24

Exploring photometric redshifts as an optimization problem: an ensemble MCMC and simulated annealing-driven template-fitting approach

Joshua S. Speagle,[1,2,3]★ Peter L. Capak,[3,4] Daniel J. Eisenstein,[2] Daniel C. Masters[3] and Charles L. Steinhardt[3]

[1]*Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwanoha 5-1-5, Kashiwa, Chiba 277-8583, Japan*
[2]*Harvard University Department of Astronomy, 60 Garden St, MS 46, Cambridge, MA 02138, USA*
[3]*Infrared Processing and Analysis Center, California Institute of Technology, MC 100-22, 770 South Wilson Ave, Pasadena, CA 91125, USA*
[4]*Spitzer Science Center, California Institute of Technology, Pasadena, CA 91125, USA*

Accepted 2016 June 21. Received 2016 June 17; in original form 2015 August 10

ABSTRACT

Using a 4D grid of ∼2 million model parameters ($\Delta z = 0.005$) adapted from Cosmological Origins Survey photometric redshift (photo-z) searches, we investigate the general properties of template-based photo-z likelihood surfaces. We find these surfaces are filled with numerous local minima and large degeneracies that generally confound simplistic gradient-descent optimization schemes. We combine ensemble Markov Chain Monte Carlo sampling with simulated annealing to robustly and efficiently explore these surfaces in approximately constant time. Using a mock catalogue of 384 662 objects, we show our approach samples ∼40 times more efficiently compared to a 'brute-force' counterpart while maintaining similar levels of accuracy. Our results represent first steps towards designing template-fitting photo-z approaches limited mainly by memory constraints rather than computation time.

Key words: methods: statistical – techniques: photometric – galaxies: distances and redshifts.

1 INTRODUCTION

Future large-scale surveys such as *Euclid* (Laureijs et al. 2011), the *Wide-Field Infrared Space Telescope* (*WFIRST*; Green et al. 2012), and the Large Synoptic Survey Telescope (LSST; Ivezic et al. 2008) that seek to constrain dark energy equation-of-state using weak gravitational lensing (Albrecht et al. 2006; Bordoloi et al. 2012) will require the derivation of redshifts (z) to an enormous number of objects ($\gtrsim 10^9$). While spectroscopic redshifts (spec-z's) often are extremely precise, their cost- and time-intensive requirements necessitate the use of 'photometric redshifts' (photo-z's) derived from fitting spectral energy distributions (SEDs) taken from a combination of broad- and/or narrow-band photometry in order to measure redshifts to majority of observed objects in a feasible amount of time.

Two main approaches are currently used to derive photo-z's:

(i) *Template fitting*, which attempts to determine the set of *forward* mappings from a collection of model parameters and templates to observed colour space.

(ii) *Machine learning*, which attempts to directly determine the best *inverse* mapping from observed colour space to redshift via a training set of multiband photometry and their corresponding spec-z's.

Template-fitting-based photo-z codes in use today suffer from several modelling and computational deficiencies. Due to an insufficient understanding of the relevant parameter space, most codes rely on fitting pre-generated 'grids' of model galaxy photometry to probe corresponding regions of interest. This crude, 'brute-force' approach not only results in inefficient sampling, but also requires trade-offs in parameter resolution in order to remain computationally viable. As a result, it is generally both too slow *and* too inaccurate to meet the stringent requirements of these future dark energy surveys, even with very sophisticated implementations (Brammer, van Dokkum & Coppi 2008; Ilbert et al. 2009).

Due in part to these issues, many researchers today have turned to machine learning as a way to meet these requirements. While current advances in machine learning-based photo-z's show much promise and perform well with good spectroscopic training sets (Carrasco Kind & Brunner 2013, 2014a; Sánchez et al. 2014; Elliott et al. 2016; Hoyle 2016), the current lack of spectroscopic coverage in specific but relevant regions of colour space (Masters et al. 2015) and difficult calibration issues (Cunha et al. 2014; Newman et al. 2015) indicate that template-fitting methods will likely still play a major part in determining good photo-z estimates for these future surveys. This will particularly relevant at higher redshifts, where spectroscopic coverage is sparser and more systematically biased.

★E-mail: jspeagle@cfa.harvard.edu

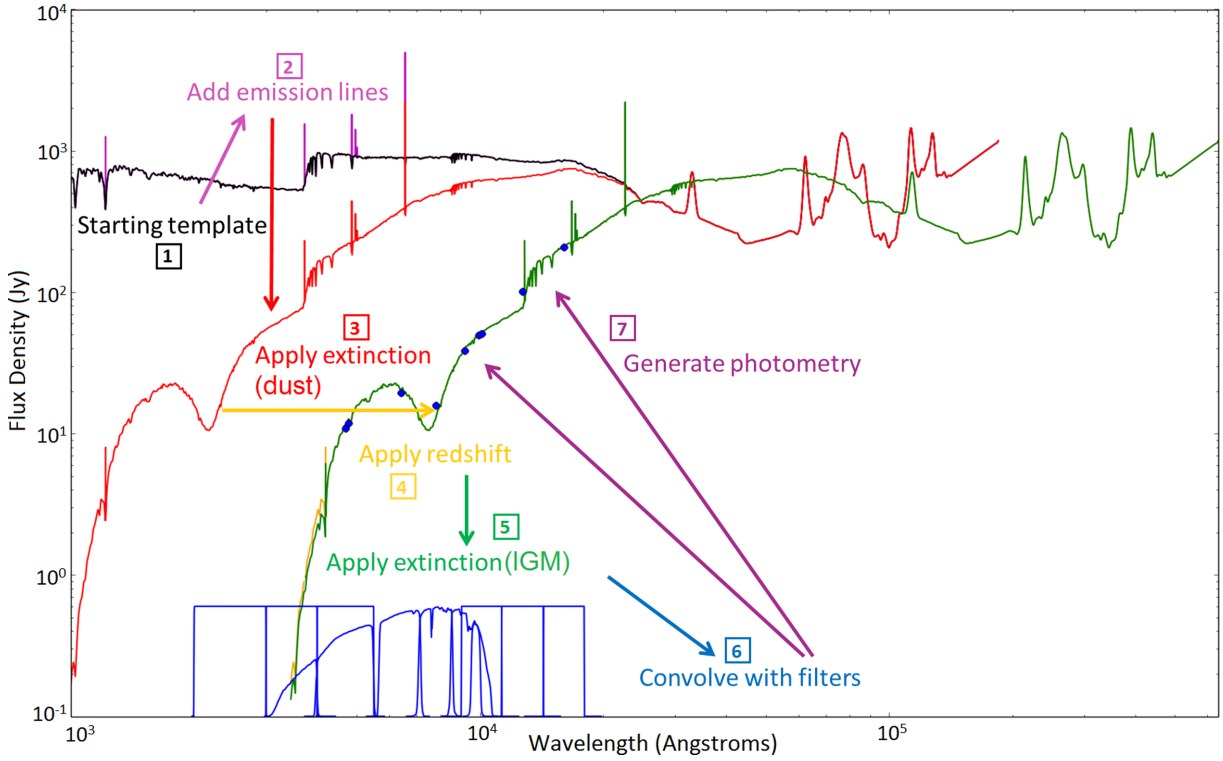

Figure 1. A schematic of how model photometry is generated for a given set of parameters and a collection of galaxy, emission line, and reddening templates. The initial galaxy template (black) is first modified by adding on emission lines (purple) before being reddened by a uniform galactic dust screen (red). The template is then redshifted (orange) and reddening from the IGM is applied (green) before being convolved with a given filter set (blue) to compute the final model photometry (blue circles).

We attempt to address some of the computational deficiencies involved in template-fitting-based photo-z searches. Our main focus is the exploration of the general likelihood surface defined by pre-generated template grids and whether minimization techniques and clever sampling – both tailored to the general properties of the surface – can subsequently accelerate photo-z calculation and form the basis of a template-based approach limited by available memory rather than computation time.

This paper is organized as follows. In Section 2, we give a brief overview of how standard template-fitting codes generate model photometry. In Section 3, we describe the mock photometric catalog and the underlying photo-z model grid that we use for testing. In Section 4, we explore the overall shape of the multidimensional likelihood surface for individual objects. In Section 5, we outline, characterize, and test a combined Markov Chain Monte Carlo (MCMC) and simulated annealing approach to sampling photo-z likelihood surfaces in a rapid and efficient manner. We discuss possible future extensions of this work in Section 6.

We standardize to the AB magnitude system (Oke & Gunn 1983) throughout the paper.

2 FROM OBSERVED SED TO PHOTO-Z

Deriving photo-z's from observed photometry is composed of three main parts.

(i) Generate model photometry from a list of input parameters $\boldsymbol{F}_{\rm model}(\boldsymbol{x})$.

(ii) Determine the goodness-of-fit (GOF) between the model photometry and the observed data $\boldsymbol{F}_{\rm obs}$.

(iii) Use this information to decide what model parameters to sample.

We discuss each of these in turn.

2.1 Generating model photometry

To generate model photometry, most codes begin with a set of 'basis' galaxy templates which are used either individually (Ilbert et al. 2009) or in some linear combination scheme (Blanton & Roweis 2007) to create an underlying galaxy template set. To incorporate the impact of specific emission lines, these galaxy templates are often modified with a set of emission line templates co-added according to a set of scaling relations taken from the literature (Ilbert et al. 2009; Salmon et al. 2015). Rest-frame galaxy templates are created by superimposing an additional *uniform screen* of galactic dust taken from a basis set of normalized dust attenuation curves (see e.g. Bolzonella, Miralles & Pelló 2000; Charlot & Fall 2000) with a given amount of extinction, usually parametrized by $E(B-V)$.

These rest-frame galaxy templates are then redshifted by $(1+z)$ and modified by extinction from the intergalactic medium (IGM; Madau 1995) to form the final observed-frame galaxy template. The corresponding set of model photometry is then generated by convolving the model galaxy flux density with the transmission of a particular filter (including atmospheric effects) normalized to a source at constant flux density. An illustration of this forward-mapping process is shown in Fig. 1.

Generating a new set of model photometry is computationally expensive, involving multiple addition, multiplication, and power/exponential operations on several large ($\gtrsim 10^4$–10^5 element)

arrays. Furthermore, although generating a new model template is around an order of magnitude or more computationally taxing than the corresponding filter convolution, the latter process on its own is still too much of a computational burden to be called frequently when computing photo-z's for individual objects.

As a result, often the only computationally feasible approach is to generate a large grid of model photometry for a discrete set of parameters *before* – rather than during – the actual SED fitting process. Although other alternatives exist (Graff et al. 2012; Akeret et al. 2015), we focus on optimizing $P(z)$ in the approximately instantaneous likelihood limit (i.e. on a pre-generated grid of model photometry) rather than in the expensive likelihood limit (where the computational overhead involved with choosing samples is much smaller than time necessary to compute the likelihood) due to our emphasis on computational speed.

2.2 Determining the GOF

In order to derive accurate multidimensional likelihoods, we must choose a suitable GOF metric. Although there are many possible choices, most routines use the simple χ^2 metric in order to incorporate uncertainties on the photometry and/or model in a straightforward manner, where

$$\chi^2(\boldsymbol{x}, s) \equiv \sum_i \left(\frac{F_{\mathrm{obs},i} - s F_{\mathrm{model},i}}{\sigma_i} \right)^2 \quad (1)$$

where s is an associated model scalefactor, $\sigma_i^2 = \sigma_{\mathrm{obs},i}^2 + \sigma_{\mathrm{model},i}^2$ is the total variance, and the sum over i is taken over all observed bands. For a given $\boldsymbol{x}$ we can marginalize over s to minimize $\chi^2(\boldsymbol{x})$, giving us

$$s = \sum_i \frac{F_{\mathrm{obs,i}} F_{\mathrm{model,i}}}{\sigma_i^2} \Bigg/ \sum_i \frac{F_{\mathrm{model,i}}^2}{\sigma_i^2}. \quad (2)$$

This is a simple one-step process that can be calculated prior to computing the actual χ^2 value.[1]

2.3 SED fitting

As outlined in Section 2.1, template-fitting methods generate new model photometry by performing a series of operations on a set of templates before convolving the final product with the relevant filter set. As this process is expensive, most current approaches choose to pre-generate a large grid of model templates containing $\sim 10^{6\text{–}7}$ individual sets of photometry. Due to the enormous reduction in computing time that can be achieved on a pre-computed grid – calling χ^2 given $\boldsymbol{F}_{\mathrm{model}}(\boldsymbol{x})$ is several orders of magnitude faster than generating a new model from scratch – pre-generating photometry from a large set of parameter combinations often saves orders of magnitude of computation time compared to repeatedly computing them in real time, especially given the large number of objects usually involved.

[1] While χ^2 is valid for normally distributed data at any level of signal-to-noise (S/N), it *does not* incorporate any method of treating upper limits, which are relatively common in astronomical data. As a result, many codes (incorrectly) implement ad hoc procedures in an attempt to include this more limited set of information, leading to small but *non-negligible* biases (Sawicki 2012). As a result, it is *much* more useful to undertake SED fitting in flux space, which can accommodate negative fluxes present in low-S/N data, rather than in magnitude space, which cannot.

The approach taken by most template-fitting photo-z codes today (e.g. LE_PHARE; Arnouts et al. 1999; Ilbert et al. 2006) is to fit the *entire* pre-generated grid of points to each object to build up a model of the full N-dimensional likelihood $P(\boldsymbol{x}|\boldsymbol{F}_{\mathrm{obs}})$ at a pre-determined resolution. Afterwards, marginalized probability distribution functions (PDFs) can be created through

$$P(\boldsymbol{x}_I|\boldsymbol{F}_{\mathrm{obs}}) \propto \int P(\boldsymbol{x}|\boldsymbol{F}_{\mathrm{obs}}) d\boldsymbol{x}_J \approx \sum_{\boldsymbol{x}_J} P(\boldsymbol{x}|\boldsymbol{F}_{\mathrm{obs}}), \quad (3)$$

where $\boldsymbol{x}_I$ are the subset of parameters of interest and $\boldsymbol{x}_J$ are the subset of parameters to be marginalized over.

While generally effective (Hildebrandt et al. 2010; Dahlen et al. 2013), these *brute-force* approaches are subject to two major issues.

(i) *Inefficient at probing region(s) of interest.* Brute-force methods tend to spend the majority of time for any given object ($\gtrsim$99 per cent) sampling regions of extremely low probability.

(ii) *Scale proportional to dimensionality of problem.* The corresponding size of the grid increases multiplicatively with the number of dimensions to be probed and the desired granularity of each dimension.

We will return to these issues in Section 5.

3 GENERATING A REALISTIC MOCK CATALOGUE

To avoid possible issues caused by template mismatch (especially when compared to the full range of observed galaxy SEDs, including emission line variation) and other possible systematic effects, we opt to explore the parameter space spanned by a mock catalogue of galaxies using the same set of templates we hope to later test.

We construct a mock catalogue using the high-quality photometry ($\sim$30 bands) available in the Cosmological Origins Survey (COSMOS; Scoville et al. 2007) field (Capak et al. 2007). Photo-z's are derived with the template-fitting photo-z code LE_PHARE using a grid with the following specifications.

(i) $N_z = 601$: spans the redshift range $z = [0, 6]$ in steps of $\Delta z = 0.01$.

(ii) $N_{\mathrm{T,gal}} = 31$: includes eight elliptical templates and 11 spiral templates based on Polletta et al. (2007), supplemented with 12 starburst (SB) templates constructed from Bruzual & Charlot (2003) stellar population synthesis models assuming exponentially declining star formation histories with ages ranging from 3 to 0.03 Gyr.

(iii) $N_{\mathrm{emline}} = 3$: only a single template including Ly α, O[II], Hβ, O[III], and Hα is used. This is added to each template prior to applying reddening effects according to $\{0.5, 1.0, 2.0\}$ times the scaling relations outlined in Ilbert et al. (2009).

(iv) $N_{\mathrm{T,dust}} = 5$: the dust curves used include observations from the SMC (Prevot et al. 1984), LMC (Fitzpatrick 1986), MW (Seaton 1979; Allen 1976), and SB galaxies (Calzetti et al. 2000).

(v) $N_{E(B-V)} = 9$: each template is allowed $E(B-V)$ values of $\{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5\}$.

After correcting for zero-point offsets using the spectroscopic sample from zCOSMOS (Lilly et al. 2007, 2009), a subset of the grid is then fit to each object and marginalized over to derive the associated redshift PDF. See Ilbert et al. (2009) for additional information.

Using the median $P(z)$ value (interpolated to $\Delta z = 10^{-4}$ resolution), best-fitting template, reddening law, $E(B-V)$, and scaling factor derived from LE_PHARE for each galaxy, we create a collection of corresponding model galaxy templates that mimic the

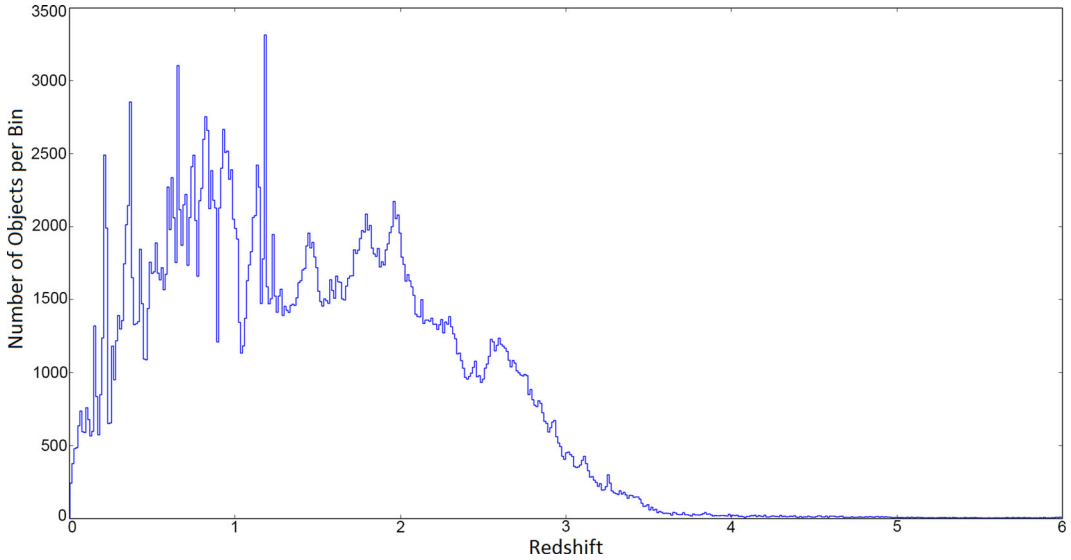

Figure 2. The redshift distribution of our mock catalogue of 384 662 COSMOS galaxies. Although the distribution extends out to $z = 6$, the majority of galaxies are located at $z \lesssim 3.2$.

observational data.[2] We convolve these templates with a set of 12 filters designed to mimic the wavelength ranges probed by the future Cosmological Advanced Survey Telescope for Optical and UV Research (CASTOR; $UV_c u_c g_c$) and *Euclid* ($Y_w J_w H_w$) missions supplemented with ground-based photometry (*ugrizY*). These provide a wide wavelength range to detect spectral features, including three overlapping bands (u, g, and Y).

From an initial sample of ∼2 million galaxies, we implement an $H_w = 24.5$ mag cut to mimic the 5σ imaging depth of the wide-field *Euclid* survey. This leaves us with ∼20 per cent of the original sample, or 384 662 galaxies spanning a redshift range from $z = 0$ to ∼3.2, including a small number of objects up to $z \sim 6$ (Fig. 2).

These photometric fluxes are then jittered according to the expected background noise levels based on the anticipated depth of the imaging in each band[3] to create the final mock catalogue. To avoid complications arising from the inclusion of upper limits (see footnote 1), we leave the model photometry in flux space.

In addition, we also compute a baseline GOF value for each object ($\chi^2_{\rm base}$) by determining the associated χ^2 values between the original (pre-jitter) model photometry and final (post-jitter) mock photometry. This allows us to see the extent to which our errors altered the original GOF while also providing a check on the overall quality of the best fits determined by our code(s). Given optimal performance and an ideal model set, the derived $\chi^2_{\rm obj}$ for an object should always satisfy the condition $\chi^2_{\rm obj} \leq \chi^2_{\rm base}$.

In other words, an ideal photo-z code should always find either (1) the 'correct' solution or (2) an 'incorrect' one that is a better fit to the data. This is not always true in practice, especially in our case where the associated redshift has been regenerated to a higher resolution than the corresponding Δz spacing of the grid. While being able to find the 'best' match to the data is a desirable feature of an effective fitting routine, it is not strictly necessary – as long as the general region *around* the best fit is probed sufficiently well (i.e. the relative *shape* of the PDF can be recovered even when sparsely sampled), a given algorithm should still be able to derive accurate $P(z)$'s and associated estimates (see Section 5).

[2] Due to the coarse nature of the emission line grid, we do not include emission line contributions to simplify the nature of our tests. This does not impact our conclusions, although we explore ways of incorporating this additional complexity in Speagle & Eisenstein (2015).

[3] This calculation does not include error from shot noise from galaxy photons, which is expected to be of the order of ∼2 per cent at the 3σ detection limit.

4 MAPPING THE PHOTO-Z LIKELIHOOD SURFACE

Using a finely spaced grid of ∼2 million sets of model photometry – identical to the one described in Section 3 except with $\Delta z = 0.005$ – we calculate the χ^2 value at every trial point for a series of individual objects from our mock catalogue. We then derive the corresponding locations and likelihoods for competing minima as a function of the Euclidean distance (normalized to the grid spacing in each dimension) from the global minimum. To determine the relative 'size' of each minimum, we force each point on the grid follow the surrounding (discrete) gradient until it locates the closest corresponding minimum and record the final number of trial points occupying each local minimum.

To understand how this underlying structure corresponds to the output $P(z)$ distribution, we marginalize over all trials to derive $P(z)$. Together, these data not only inform us about the approximate general distribution, size, depth, and behaviour of minima within the 4D parameter space, but also how the corresponding structure affects the final marginalized distribution of interest. The results for three representative objects are plotted in Fig. 3.

There are several main features of interest.

(i) The main global minimum is surrounded by numerous competing minima that occupy sizeable regions of parameter space (middle panels), indicating that the space directly in the area of the global minimum and other notable degeneracies are quite 'bumpy'.

(ii) Degeneracies at moderate distances appear to occupy a significant region of parameter space (bottom panels), although its overall contribution to the marginalized likelihood is negligible here (top panels).

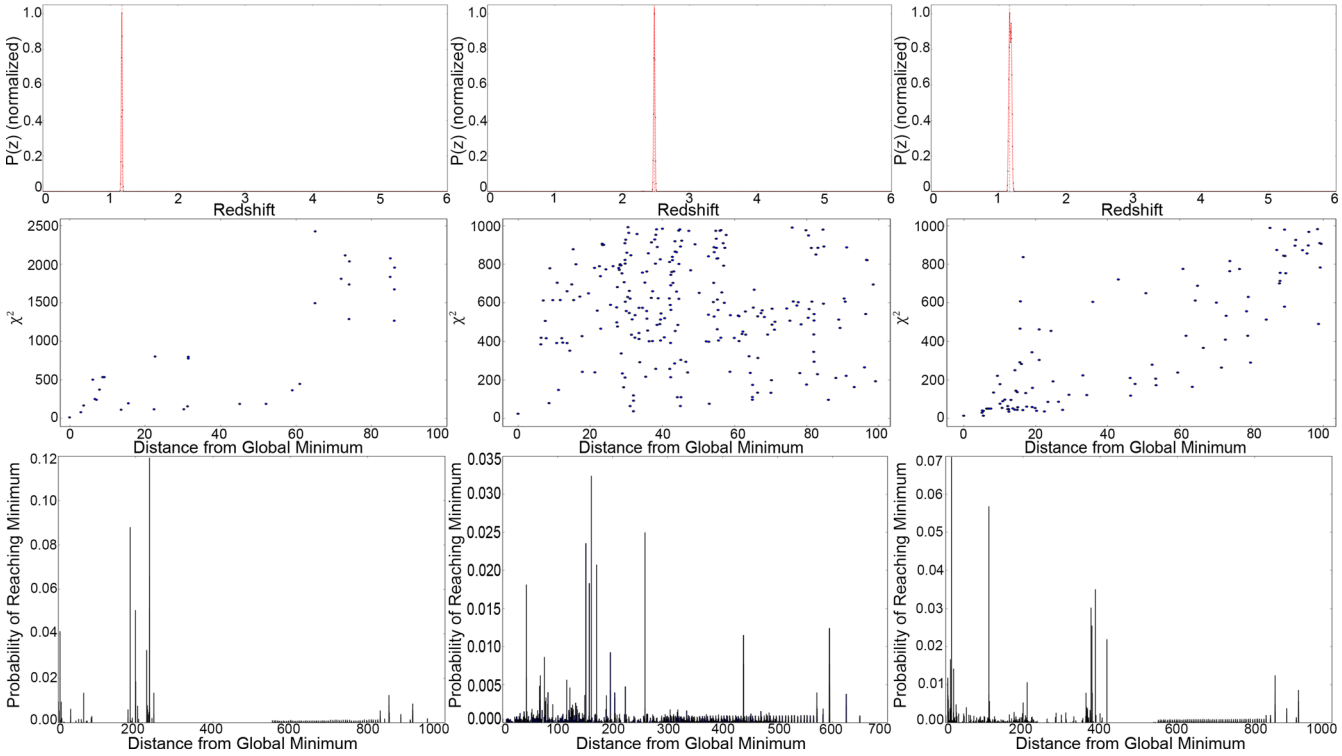

Figure 3. 4D maps of the photo-z likelihood surface for three representative objects from our mock catalog shown from left to right. Top: marginalized $P(z)$ distributions for each object (solid red) with the location of the peak marked by a dashed red line. In all cases $P(z)$ distribution is narrow and well defined. Middle: χ^2 values of individual minima within 100 Euclidean grid units ($\Delta z \sim 0.5$) from the global minimum. Although the region around the global minimum is quite 'bumpy' (i.e. filled with secondary solutions), most have a negligible impact on the marginalized $P(z)$'s (top panels). Deriving accurate $P(z)$'s thus relies heavily on being able to locate and sample the small region surrounding the global minimum. Bottom: the probability (normalized to 1) that a random trial point on the grid will reach a specific minimum as a function of distance. While the global minimum is a far better fit than most competing minima (middle panels), the latter occupy large areas of parameter space, making it extremely difficult for local and/or gradient-based minimization algorithms to reach relevant $P(z)$ modes. See Section 4 for additional details.

(iii) A series of small 'ridges' within the corresponding region of parameter space (bottom panels) that have negligible probability but occupy a small region of parameter space that needs to be avoided.

Most crucially, almost none of this substructure appears in the final redshift PDF (top panels). As a result, not only does the full 4D space appear to contain hundreds to thousands of local minima with degenerate regions a significant distance away from the global minimum occupying large regions of parameter space, but most of this structure remains hidden when investigating solely marginalized 1D $P(z)$ distributions.

In addition to creating complex surfaces in the full 4D parameter space, many of these degeneracies also are present in reduced 2D subspaces – most crucially, in the z–$E(B-V)$ 2D subspace at fixed galaxy and dust template. Using a combination of constrained gradient-descent algorithms (e.g. COBYLA; Powell 1994) and stochastic *metaheuristics* (e.g. 'basin-hopping'; Wales & Doye 1997), we find that secondary minima and elongated degeneracies are still dominate the gradient, such that ∼1500 trials (with ∼10 restarts) are necessary to reliably identify the true global minimum. This corresponds to ∼30 per cent of the size of the underlying z–$E(B-V)$ grid, albeit sampled at arbitrary (rather than fixed) resolution. While the true efficiency for an stochastically driven gradient-descent algorithm tailored to the specifics of the photo-z parameter space would likely be somewhat higher, these results confirm that the degeneracies observed in higher dimensions are also reflected in lower dimensional subspaces.

Much of this substructure highlights features and degeneracies corresponding to physical galaxy properties well known in the literature. For instance, the largest clusterings of secondary minima seem to occur around 100 – 200 grid units ($\Delta z = 0.5 - 1$) away in Fig. 3 represent the redshift-reddening degeneracy, whereby intrinsically redder objects are confused with bluer objects observed at higher redshift (and vice versa). This degeneracy is often represented multiple times (governing both low-z or high-z solutions from a variety of template/dust combinations), with one to two solutions generally favoured.

In addition, a cluster of fits at very large distances from the global minimum is the result of confusion between the 1216 Å Lyman and ∼4000 Å Balmer breaks (i.e. the high-low redshift degeneracy). These are only represented once given the limited redshift range spanned by our grid and objects, but could appear multiple times if the redshift range is broadened significantly.[4] Their lower significance for our objects is due to our colour-space coverage, especially in the near-IR (from the *Euclid YJH* bands) and near-UV (from the CASTOR UV and *u* bands).

Not all features seen in Fig. 3 correspond to physical properties, however – many are instead artefacts of the grid used during the fitting process. For instance, each of the galaxies displayed in Fig. 3 contain prominent attractor solutions located at $z = 0$ ($d \sim z/0.005$),

[4] Assuming no priors are instituted preventing against almost certainly unphysical solutions.

whose amplitudes depends on the galaxy's specific SED (which governs the gradient). In addition, many of the 'furrows' seen in the bottom panels are caused by edge effects that substantially increases the number of local minima present and whose amplitude again depends on the specific galaxy SED (see e.g. the central object in Fig. 3).

Results from Speagle & Eisenstein (2015) suggest that much of this substructure is due to the combination of 1D projections of non-linear effects in the creation of the grid. Galaxy SEDs occupy a very non-linear N-dimensional manifold in colour space, with complex and correlated changes in the overall shape. However, when used in most grids, they are often assigned arbitrary numbers based on some simple diagnostics (e.g. FUV flux), equivalent to a crude 1D projection. This projection loses valuable information concerning the higher dimensional manifold, and when combined with the equivalently crude projection of dust attenuation curves (where the ordering is often equally arbitrary) and redshift evolution, creates a number of (wide) local minima that might otherwise not exist. While this indicates that a more suitable choice of grid might eliminate some of the bumpiness and/or edge effects, the broad features observed in Fig. 3 still remain valid in general.

5 DESIGNING AN EFFICIENT AND ROBUST PHOTO-z SEARCH ALGORITHM

To efficiently sample the complex and bumpy likelihood surface described in Section 4 without resorting to brute-force methods (see Section 2.3), we explore a combination of MCMC sampling and simulated annealing. This allows us to sample high-dimensional grids efficiently at arbitrary resolution and perform well in bumpy, multimodal parameter spaces. We describe each portion of our approach below.

5.1 MCMC

Unlike brute-force approaches, which sample evenly and apply weights afterwards, MCMC-based algorithms (see e.g. `SatMC`; Johnson et al. 2013) sample at a rate proportional to the PDF itself and weight every *accepted* trial point *evenly* (although see Bernton et al. 2015). A standard search heuristic employed by most MCMC codes is the Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970).

(i) Draw a set of trial parameters $\boldsymbol{x}_n$ from the neighbourhood function[5] $q(\boldsymbol{x}|\boldsymbol{x}_{n-1})$.

(ii) Accept the new trial and move to location $\boldsymbol{x}_n$ with probability $\min\left(1, \frac{P(\boldsymbol{x}_n|\boldsymbol{F})}{P(\boldsymbol{x}_{n-1}|\boldsymbol{F})}\frac{q(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})}{q(\boldsymbol{x}_{n-1}|\boldsymbol{x}_n)}\right)$. Otherwise, remain at $\boldsymbol{x}_{n-1}$.

(iii) Repeat from step (i) until a stopping criterion is reached.

This procedure is used to guide several individual 'chains' of related draws as they converge ('burn in') to and eventually begin sampling from the region of interest. Although the exact neighbourhood function chosen is ultimately arbitrary, most often an N-dimensional multivariate-normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is used, where $\boldsymbol{\mu} = \boldsymbol{x}_{n-1}$ is the mean vector (adjusted at each step) and $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ is the covariance matrix ($\mathbf{I}$ being the identity matrix), also often adjusted over the course of a typical run.

Because MCMC-based algorithms sample at a rate approximately proportional to the PDF in a given region of interest, they are able to explore large, N-dimensional spaces at finer resolution and with far fewer function calls than grid-based approaches while scaling relatively slowly with the dimensionality of the problem (up to moderate dimensionality; Handley, Hobson & Lasenby 2015). As a result, An MCMC-based approach running on a large, pre-generated grid represents a more efficient way of developing a model of the likelihood by exploring only the portion of the grid that contains the majority of the probability (see e.g. `SpeedyMC`; Acquaviva, Gawiser & Guaita 2012). As a result, *arbitrarily large grids can be used without substantially increasing the computation time*, provided they can be loaded into memory. This enables us to increase the resolution and range of parameter space we are allowed to explore, such as by expanding the number of galaxy (Brown et al. 2014) and dust templates (Fitzpatrick & Massa 2007).

For our particular purposes, we use the ensemble MCMC sampling implementation from Goodman & Weare (2010) and implemented in EMCEE[6] (Foreman-Mackey et al. 2013) as the underlying foundation of our algorithm. In brief, this approach replaces the above chains with a (much larger) ensemble of 'walkers' that utilize their collective distribution to propose new trials.[7] See Foreman-Mackey et al. (2013) for additional details of the benefits (and drawbacks) of this approach (although see Huijser, Goodman & Brewer 2015).

5.2 Simulated annealing

Although MCMC is effective at sampling around the region of interest, due to the complex nature of the likelihood surface it will likely encounter substantial difficulties locating the global minimum during the burn-in phase. To assist with this, we turn to *simulated annealing*, a metaheuristic designed to assist searches such as these where the goal is to find the global minimum in a bumpy and often significantly multimodal space.

Simulated annealing involves imposing a global temperature on either the entire space (the standard implementation) or individual samplers/regions (otherwise known as 'tempering'; see Johnson et al. 2013) that distorts the shape of the space such that

$$P(\boldsymbol{x}) \rightarrow [P(\boldsymbol{x})]^{T_0/T(t)}, \tag{4}$$

where $T(t)$ is the temperature as a function of time (i.e. iteration) and T_0 is the 'transition temperature' that we henceforth take to be 1. For $T(t) > T_0$, 'bad' jumps have a higher relative probability of being accepted, allowing an algorithm additional stochasticity while sampling. As a result, it is able to explore more of the search space in the hopes of finding the region around the global minimum. For $T(t) < T_0$, bad jumps instead become relatively *less* likely of being accepted, reducing stochasticity and increasingly constraining the algorithm to only move in the direction of the gradient. Ideally, this forces our algorithm to find a more optimal solution assuming that it has reached the general region of interest.

5.3 Implementation

While simulated annealing is effective at locating the general area of the peak, because it distorts the likelihood over the course of the run (thus violating 'detailed balance') samples cannot be used to reconstruct the associated PDF. As a result, we only utilize simulated

[5] Also often referred to as a 'proposal distribution'.

[6] http://dan.iel.fm/emcee

[7] We have chosen to keep the associated scalefactor $a = 2$ throughout this paper, since like Foreman-Mackey et al. (2013) we find no evidence it impairs performance.

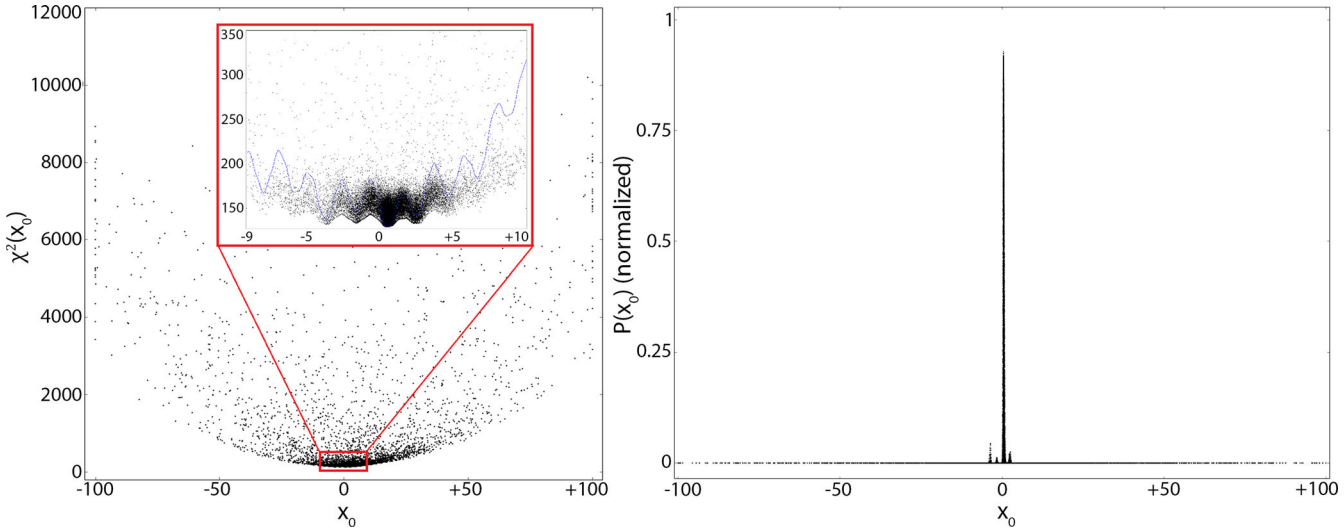

Figure 4. A 1D projection of the χ^2 (left) and likelihoods $P = e^{-\frac{\chi^2}{2}}$ (right) of all accepted trials generated from a toy 4D function designed to broadly mimic the features of our photo-z space taken from a fiducial run of our proposed MCMC-based, simulated annealing-driven algorithm. A portion of a 1D component of this function is shown in blue (left central inset). The dense sampling in regions surrounding the global minimum illustrates that the algorithm is sampling effectively and can robustly locate regions of high probability in significantly bumpy parameter spaces.

annealing during the burn-in phase to help locate the global minimum. Afterwards, we set $T(t)$ to 1, re-spawn an ensemble of walkers in an N-dimensional multivariate Gaussian distribution around the best-fitting value located by the ensemble[8], and re-construct $P(z)$ from a subset of the latter trials.

In summary, our implementation is as follows.

(i) Initialize W walkers on an arbitrary N-dimensional input grid drawn from a uniform distribution.

(ii) Start a simulated annealing-driven ensemble MCMC run with a given set of $T(t = 0)$ and $\Delta T/\Delta$(ensemblerun) values to find the region of the global minimum.

(iii) Spawn a new distribution of W walkers in an N-dimensional multivariate Gaussian with some fractional spread $\sigma_{\rm frac}$ in each dimension around the trial with the highest likelihood from the previous simulated annealing-driven run.

(iv) Set $T(t) = 1$ and $\Delta T/\Delta$(ensemblerun) $= 0$ and start a new run for an additional $N_{\rm MCMC}$ number of ensemble runs.

(v) After discarding an initial fraction $f_{\rm disc}$ of trials from this second run, reconstruct the redshift PDF using the remaining accepted trials.

5.4 Performance on a toy function

Before applying our algorithm to real data, we first characterize its performance on a sample function designed to mimic the overall properties of the photo-z parameter space we observed in Section 4. We define the log-likelihood (i.e. χ^2) of this function as the product of four 1D underlying quadratic likelihoods with numerous sin/cosine 'bumps' superimposed with varying periods (widths) and depths (amplitudes). This gives our function an overarching gradient filled with numerous minima, a somewhat idealized case of the behaviour observed in Section 4. A portion of this function is plotted in the inset of Fig. 4 to give a rough sense of the overall shape.

We find that the performance of our combined sampling approach on our sample function is relatively insensitive to the number of walkers as long as they are sufficiently numerous. For a fixed annealing schedule, $N \gtrsim 150$ number of walkers is often adequate to ensure sufficient sampling of the likelihood. In addition, for a fixed number of walkers, we find convergence behaviour is generally most robust when >50 per cent of trials take place with $T > 1$, and when the annealing schedule is no faster than $\Delta T/\Delta t = 0.05$ (where t measures a single run through the full set of walkers), which ensures at least 40 ensemble runs (∼6000 individual trials) take place for $T < 1$. This gives the walkers plenty of time to converge to and eventually locate the maximum likelihood solution rather than possible secondary minima.

Once the walkers have located a potential global minimum, subsequent burn-in and sampling of the likelihood is relatively insensitive to the size of the N-dimensional multivariate Gaussian used to regenerate the distribution for values of $\sigma_{\rm frac} \gtrsim 2$ times as large as the width of the global minimum (in this case ∼0.5 per cent the parameter range spanned in each dimension). Finally, we find that, given ∼150 walkers, the collective ensemble needs ∼25 ensemble runs to reliably thermalize and begin sampling from the likelihood in an unbiased function.

These results lead us to adopt conservative parameter choices of $N = 250$ walkers, $T(t = 0) = 2.5$, $\Delta T/\Delta t = 0.01$, $\sigma_{\rm frac} = 0.05$, $N_{\rm MCMC} = 100$, and $f_{\rm disc} = 0.4$ to characterize the fiducial behaviour of our combined MCMC and simulated annealing approach. Using these parameters, we verify that not only is the combined algorithm able to effectively locate and characterize the region of interest for our toy function, but that the performance at both low and high T is indeed important in locating the global minimum, effectively burning in to the region around the global minimum during the $T > 1$ period while centring in on the maximum likelihood value during the $T < 1$ phase. A rough illustration of our algorithm's overall performance is shown in Fig. 4.

[8] While re-spawning the ensemble of walkers around the best-fitting set of parameters will by default miss truly multimodal $P(z)$'s, we have chosen this approach as a compromise to give a slightly biased view of the 'true' $P(z)$ rather than completely undersampling and/or mischaracterizing it.

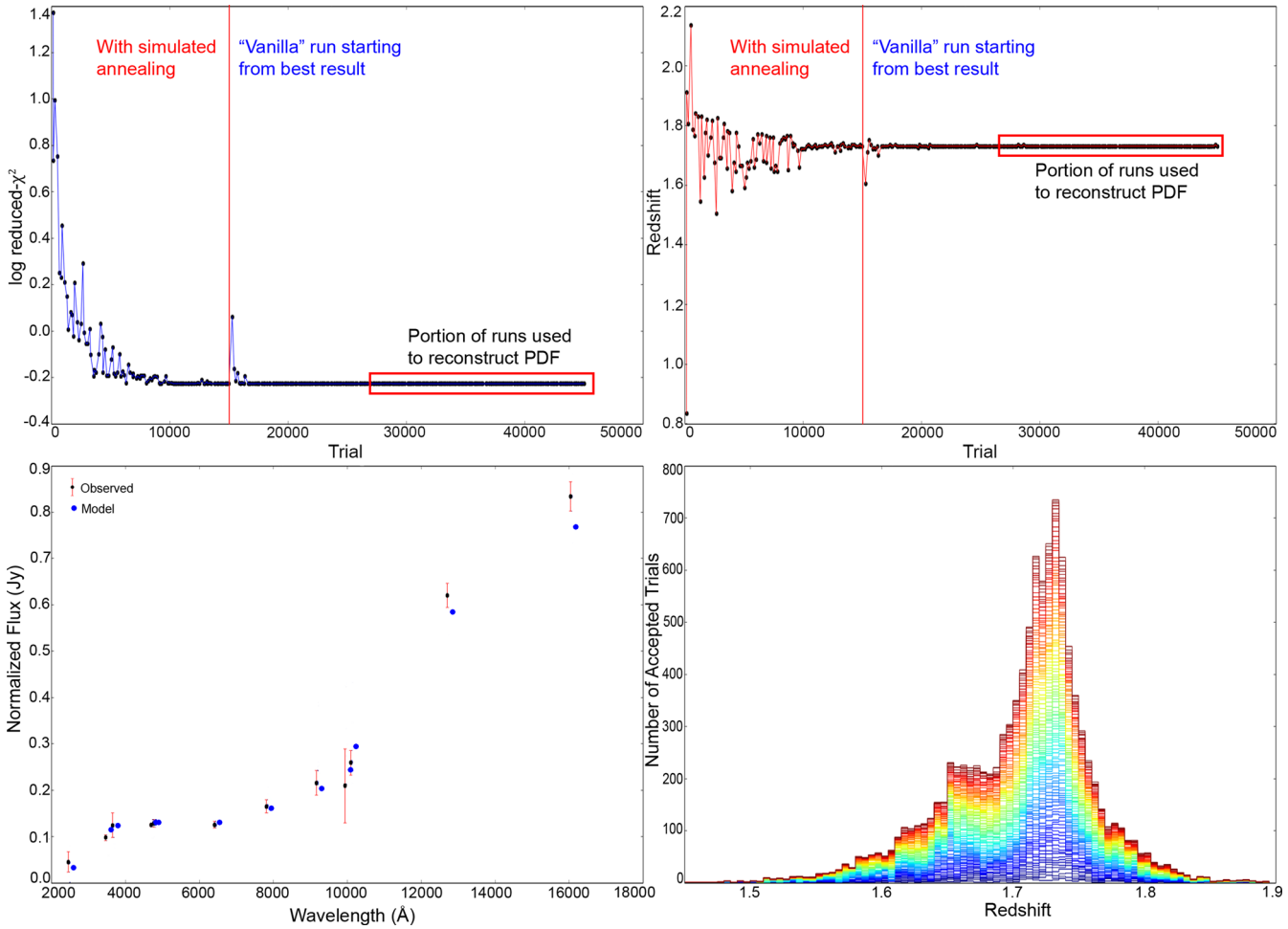

Figure 5. Top left: the minimum log (reduced-χ^2) found among the full set of walkers (see Section 5.1) after each ensemble run for a representative object in our mock catalogue. In this specific case, we converge to the best-fitting set of parameters $\boldsymbol{x}_{\rm best}$ after ∼60 per cent of the simulated-annealing driven portion of the run (0.6 per cent the size of the corresponding grid), and relocate it relatively quickly after the walkers have re-spawned for the 'vanilla' ($T(t) = 1$) portion. Top right: as the top-left panel, but for the best-fitting redshift $z_{\rm best}$. Similar behaviour can be seen, although small perturbations away from $z_{\rm best}$ are now visible. Bottom left: the best-fitting model (blue circles) versus observed (black circles with red error bars) photometry plotted at effective wavelength in each filter. For clarity, the model photometry has been slightly shifted. The observed agreement between the two sets of points indicates the overall quality of the final fit(s) as seen in the top-left panel. Bottom right: $P(z)$ derived from the final 40 per cent of all trials as a function of time (blue to red). As we are consistently probing the area directly around $\boldsymbol{x}_{\rm best}$ (see top left), MCMC sampling is able to effectively reconstruct the underlying PDF using only 18 000 independent trials, or about 1.1 per cent of the size of the corresponding grid.

5.5 Optimization using mock data

Based on results from our toy function, we now seek to characterize the performance of our algorithm on our mock data. Using a few dozen representative objects, we find that the set of fiducial parameters established in Section 5.4 give robust and acceptable results for the majority of mock data, although we find increasing the fraction of time spent at $T > 0$ from Section 5.4 increases the robustness of the burn-in process. This is likely due to the larger impact of broader but shallower likelihoods spanned by degenerate redshift solutions on the multidimensional photo-z likelihood surface (see Section 4).

Based on trial and error, we find that a fiducial set of parameters $N_{\rm walkers} = 150$, $T(t = 0) = 3.0$, $\Delta T/\Delta t = 0.03$, $\sigma_{\rm frac} = 0.15$, $N_{\rm MCMC} = 200$, and $f_{\rm disc} = 0.4$ is a reasonable compromise between optimizing runtime performance and robust recovery of photo-z PDFs. Note that this is a deliberately *conservative* choice of parameters that leads to a total of 45 150 trials (2.7 per cent of the full grid) per object (i.e. 301 ensemble runs), of which only 18 000 (1.1 per cent of the full grid) are used in the reconstruction of the final PDF. A representative run for an individual object is shown in Fig. 5.

5.6 Comparison with brute-force approach

In order to showcase an appropriate baseline to characterize the relative performance of our approach, we create a brute-force counterpart that functions exactly as described in Section 2.3, a process that involves a total of $\sim 1.6 \times 10^6$ trials (compared with 4.5×10^4 trials) per object. We run both codes on the mock photometric catalogue described in Section 3 using the same grid outlined in Section 4.

Using the reconstructed $P(z)$ for each individual object, we classify the best-fitting redshift as the median of the distribution $z_{\rm med} = {\rm median}(P(z))$. The 2D distributions of input redshift $z_{\rm in}$ versus fitted redshift $z_{\rm fit} = z_{\rm med}$ for both methods are shown in Fig. 6 up to $z = 3.2$. Although our new approach samples ∼40 times

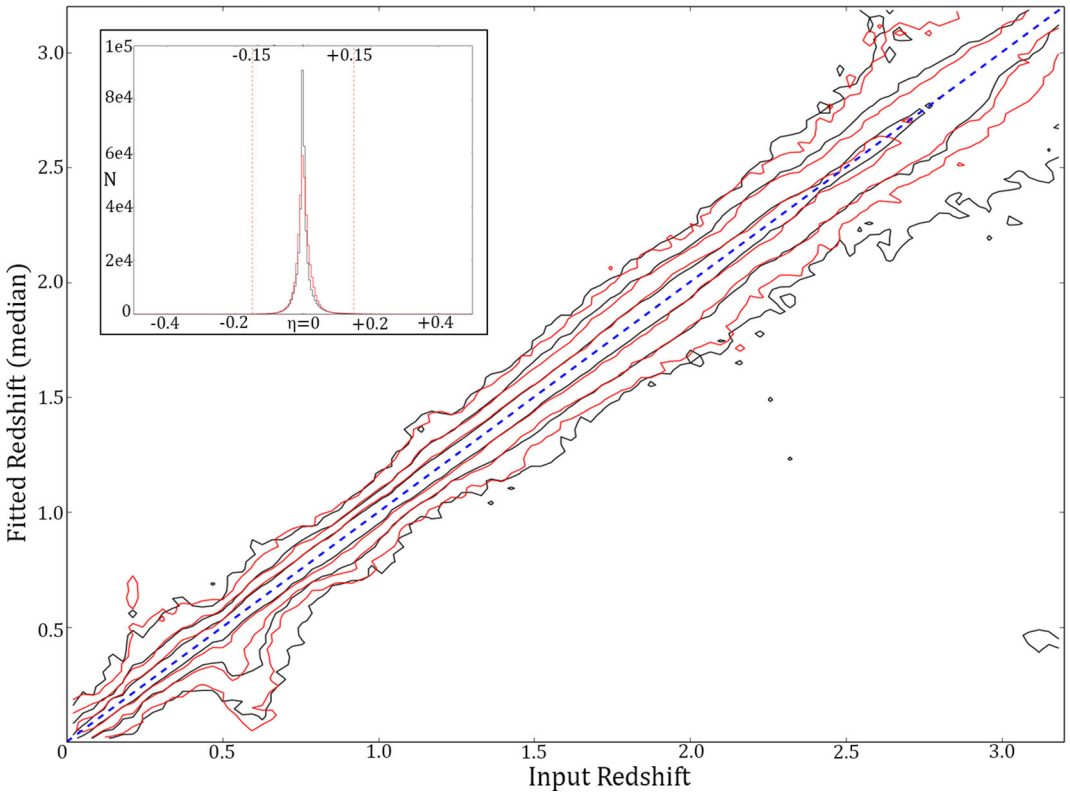

Figure 6. The distribution of input redshifts $z_{\rm in}$ versus fitted redshifts $z_{\rm fit}$ (70, 95, and 99 per cent contours plotted) from our catalogue with $z \leq 3.2$ for a traditional brute-force approach (black) and our new MCMC-based, simulated annealing-driven approach (red). A grid of ∼2 million elements ($\Delta z = 0.005$) was used. The marginalized distribution of the fractional redshift error $\eta \equiv |z_{\rm fit} - z_{\rm in}|/(1 + z_{\rm in})$ is shown in the upper-left inset with the same colour scheme, with the often-used 15 per cent threshold for 'catastrophic outliers' indicated with dashed red lines. In addition to being ∼40 times more efficient than brute-force methods, adaptive, MCMC-based sampling provides similar levels of overall accuracy while retaining the ability to accurately capture the underlying redshift PDF.

less than its brute-force counterpart, it produces photo-z estimates with comparable accuracy to the brute-force approach and similar catastrophic outlier[9] rates (0.7 and 1.0 per cent, respectively).[10]

We find our new approach is quite effective at finding good fits to the data, locating 'optimal' ($\chi^2_{\rm fit}/\chi^2_{\rm base} \leq 1$) fits in ∼55 per cent of cases and 'reasonable' ($\chi^2_{\rm fit}/\chi^2_{\rm base} \leq 5$) ones in ∼90 per cent. As the resulting redshifts are accurate in the majority of fitted objects (Fig. 6), this indicates that even in cases where ensemble MCMC sampling fails to find the 'best' fit to the data (for a variety of possible reasons), it still manages to probe the surrounding region to high enough accuracy that the marginalized $P(z)$ distribution gives accurate predictions.

Comparing the similarity between the $P(z)$'s computed from `BAD-Z` to those computed with our brute-force code, we also can determine the general ability of our new method to recover accurate PDFs as well as the overall redshift distribution. Examining the [16th, 50th, 84th] quartiles of the differences between the photo-z solutions computed with both codes, we find deviations of [−3, −0.5, and 1.5 per cent], respectively. This indicates that while our MCMC-based approach has a slight negative bias compared to the full brute-force solution, the overall agreement between both codes is excellent.

We do find, however, that the marginalized 1D distributions of the redshift-normalized errors $(z_{\rm fit} - z_{\rm in})|/(1 + z_{\rm in})$ is slightly broader for our new approach than the brute-force solution, with a two-sample Kolmogorov–Smirnov (KS) test statistic of $p = 0.03$ (2.2σ). This mostly is due to a slight broadening near the peak of the distribution, as shown in the top-left panel of Fig. 6, and suggests the MCMC-based $P(z)$'s are slightly 'rougher' (and thus are somewhat less precise) than their brute-force counterparts.

To confirm this, we investigate $P(z)$'s of individual objects, finding that the MCMC PDFs are indeed preferentially broadened (but not offset) relative to the brute-force $P(z)$'s by ∼50 per cent (median). This is likely caused by the limited number of samples used to reconstruct the PDF (due to our conservative choice of $f_{\rm disc}$) and the discrete nature of the grid (which can lead to a 'pile-up' of our walkers at specific grid points, reducing their effective resolution). Ultimately, we find both methods produce $N(z)$ distributions that are consistent with being drawn from the same parent population as both the input redshift distribution and each other (KS p-values >0.05 in all cases), indicating that although the MCMC-based PDFs might be slightly 'rougher' than those derived from sampling the entire grid, they provide a reliable and unbiased probe of the true $P(z)$.

Finally, we test the ability of our new approach to sample larger grids in approximately constant time (as claimed in Section 5.1) by re-fitting our mock catalogue using a larger input grid involving a significantly finer/broader $E(B-V)$ grid of $E(B-V) = [0, 0.8]$, $\Delta E(B-V) = 0.01$. Compared to the original input grid of

[9] Defined as $|z_{\rm fit} - z_{\rm in}|/(1 + z_{\rm in}) > 0.15$.
[10] We find similar results using redshift estimates derived from the mean and mode of each $P(z)$, which are not shown.

$E(B-V) = \{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5\}$, this represented a ninefold increase in parameter space that we expected should not improve results due to the coarser nature of the underlying input $E(B-V)$ grid.

We found our performance among both codes to be consistent with our previous analysis, as expected. However, while the brute-force run took ∼9 times as long due to the larger size of the input grid, our simulated annealing and MCMC-based approach witnessed no significant increase in runtime. In addition, *both* methods gave rise to easily identifiable 'attractor solutions' when examining the distribution of $E(B-V)$ likelihoods on the output grid, confirming the ability of our approach to adapt to the resolution of the underlying grid without the additional increase in trials needed by brute-force approaches. This confirms the ability of our algorithm to adaptively explore arbitrarily large input parameter grids in approximately constant time, a first step towards establishing a memory-limited (rather than computationally-limited) approach to deriving photo-z's using template-fitting techniques.

In summary, by understanding the general topology of the relevant photo-z likelihood surface seen by a specific grid of pre-computed model photometry, we are able to design an algorithm that is substantially more efficient than traditional brute-force approaches while retaining similar levels of accuracy. Our ensemble MCMC-based, simulated annealing-driven approach performs extremely well on the bumpy, degenerate photo-z likelihood surface explored in this work, giving good fits in >90 per cent of cases with a comparable catastrophic outlier fraction (0.7 per cent versus 1.0 per cent).

Although these results are promising, we wish to emphasize that our approach is subject to the limitations inherent in all MCMC-based approaches regarding the stochastic reconstruction of PDFs as a function of sample size and the limitations of locating and converging to the target distribution during burn-in. While we have attempted to bypass these problems through the use of ensemble MCMC sampling (to allow for more exploration of the parameter space and better reconstruction of the PDF), simulated annealing (to help ensure robust convergence to the global minimum), and the large width of our 're-spawned' distribution after burn-in (with a 1σ width of 15 per cent of the corresponding parameter space), our chosen method by construction *will* mischaracterize truly multimodal distributions, especially those with a large number of widely separated modes with similar amplitudes.

6 CONCLUSION

While photometric redshifts (photo-z's) represent an integral part of modern extragalactic science, outstanding issues that currently plague template-fitting-based approaches are concerning in the face of looming future 'big data'-oriented surveys. This work represents the first steps towards moving template-fitting photo-z codes from a runtime-limited regime to a memory-limited one. Our main results are as follows.

(i) Using a pre-generated grid of ∼2 million elements ($\Delta z = 0.005$), we create 'maps' of the associated photo-z likelihood surfaces. For our chosen grid, we find that the surface is significantly 'bumpy' in both the full 4D parameter space and 2D subspaces, with a substantial number of minima occupying large areas compared to the region directly surrounding the global best-fitting value.

(ii) Building on these results, we design a specific algorithm to explore pre-generated grids of model photometry during photo-z searches through a combination of ensemble MCMC sampling and simulated annealing.

(iii) Using a mock catalogue of 384 662 COSMOS galaxies, we our new method's performance over a wide wavelength (UV*ugrizYJH*) and redshift ($0 < z \lesssim 3.2$) range. Compared to a brute-force counterpart, we find our MCMC-based, simulated annealing-driven approach is ∼40 times more computationally efficient, retains a similar level of accuracy, and performs robustly over the entire redshift range probed.

These results are merely the first steps towards a rigorous attempt to improve photo-z's and only part of a much larger, extended effort within the extragalactic astronomical community. For instance, almost all photo-z codes – including the one showcased in this work – utilize exclusively colour information to derive $P(z)$. This, however, ignores potentially important information such as clustering, morphology, angular size, and/or surface brightness, all of which might either improve accuracy or help distinguish the dominant mode(s) of a multimodal redshift PDF. In particular, incorporating clustering information would be a useful next step towards exhausting all available information contained in photometric surveys (Ménard et al. 2013; Aragon-Calvo et al. 2015; Newman et al. 2015).

In addition, this work has focused almost exclusively on template-fitting approaches to deriving photo-z's, bypassing a wide range of machine learning techniques that are almost certainly critical in improving upon current photo-z methodologies. In particular, (un)supervised machine learning approaches such as Self-Organizing Maps (Kohonen 1982; Kohonen 2001; Carrasco Kind & Brunner 2014a; Masters et al. 2015) offer the opportunity to move beyond simple inverse mapping approaches to instead incorporate prior knowledge about a given data set in increasingly sophisticated ways (Dahlen et al. 2013; Carrasco Kind & Brunner 2014b; Kim, Brunner & Carrasco Kind 2015).

Finally, while this work has focused on many of the more computationally oriented avenues towards improving photometric redshifts, a major unresolved issue in current photo-z searches is the dual set of model uncertainties that arise due to the use of local galaxy templates (and emission line scaling relations) to probe galaxies at much higher redshifts and the limited range of dust templates (and lack of priors; Repp, Ebeling & Richard 2016) often used during the fitting process. Both of these avenues must be explored further in order to develop an improved set of templates and a better understanding of modelling uncertainties. Ultimately, there remain ample opportunities for further investigation in advancing new techniques, creating superior template sets, and improving overall computational efficiency.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referee for useful comments that substantially improved the quality of this work. JSS would like to thank Douglas Finkbeiner and Zachary Slepian for comments that improved the quality of this work and Charles Alcock for supervising the senior thesis course where much of this work was completed. JSS is grateful for financial support from the Herchel Smith-Harvard Summer Undergraduate Research Fellowship, the Harvard University Department of Astronomy, the Harvard College Observatory, and CREST funding from the Japan Science and Technology Agency (JST). This work has benefited extensively from access to computing resources at IPAC and Harvard University.

REFERENCES

Acquaviva V., Gawiser E., Guaita L., 2012, in Tuffs R. J., Popescu C. C., eds, Proc. IAU Symp. 284, The Spectral Energy Distribution of Galaxies. Cambridge Univ. Press, Cambridge, p. 42

Akeret J., Refregier A., Amara A., Seehars S., Hasner C., 2015, J. Cosmol. Astropart. Phys., 08, 043

Albrecht A. et al., 2006, preprint (astro-ph/0609591)

Allen D. A., 1976, MNRAS, 174, 29p

Aragon-Calvo M. A., Weygaert R. v. d., Jones B. J. T., Mobasher B., 2015, MNRAS, 454, 463

Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540

Bernton E., Yang S., Chen Y., Shephard N., Liu J. S., 2015, preprint (arXiv:1506.08852)

Blanton M. R., Roweis S., 2007, AJ, 133, 734

Bolzonella M., Miralles J.-M., Pelló R., 2000, A&A, 363, 476

Bordoloi R. et al., 2012, MNRAS, 421, 1671

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Brown M. J. I. et al., 2014, ApJS, 212, 18

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Capak P. et al., 2007, ApJS, 172, 99

Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483

Carrasco Kind M., Brunner R. J., 2014a, MNRAS, 438, 3409

Carrasco Kind M., Brunner R. J., 2014b, MNRAS, 442, 3380

Charlot S., Fall S. M., 2000, ApJ, 539, 718

Cunha C. E., Huterer D., Lin H., Busha M. T., Wechsler R. H., 2014, MNRAS, 444, 129

Dahlen T. et al., 2013, ApJ, 775, 93

Elliott J., de Souza R. S., Krone-Martins A., Cameron E., Ishida E. E. O., Hilbe J., 2016, Proc. Astrophysics and Space Science, Vol. 42. The Universe of Digital Sky Surveys. Springer, Switzerland, p. 91

Fitzpatrick E. L., 1986, AJ, 92, 1068

Fitzpatrick E. L., Massa D., 2007, ApJ, 663, 320

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65

Graff P., Feroz F., Hobson M. P., Lasenby A., 2012, MNRAS, 421, 169

Green J. et al., 2012, preprint (arXiv:1208.4012)

Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 450, L61

Hastings W., 1970, Biometrika, 57, 97

Hildebrandt H. et al., 2010, A&A, 523, A31

Hoyle B., 2016, Astron. Comput., 16, 34

Huijser D., Goodman J., Brewer B. J., 2015, preprint (arXiv:1509.02230)

Ilbert O. et al., 2006, A&A, 457, 841

Ilbert O. et al., 2009, ApJ, 690, 1236

Ivezic Z. et al., 2008, preprint (arXiv:0805.2366)

Johnson S. P., Wilson G. W., Tang Y., Scott K. S., 2013, MNRAS, 436, 2535

Kim E. J., Brunner R. J., Carrasco Kind M., 2015, MNRAS, 453, 507

Kohonen T., 1982, Biol. Cybern., 43, 59

Kohonen T., 2001, Self-Organizing Maps. Springer-Verlag, Berlin, p. 501

Laureijs R. et al., 2011, preprint (arXiv:1110.3193)

Lilly S. J. et al., 2007, ApJS, 172, 70

Lilly S. J. et al., 2009, ApJS, 184, 218

Madau P., 1995, ApJ, 441, 18

Masters D. et al., 2015, ApJ, 813, 53

Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, preprint (arXiv:1303.4722)

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, J. Chem. Phys., 21, 1087

Newman J. A. et al., 2015, Astropart. Phys., 63, 81

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Polletta M. et al., 2007, ApJ, 663, 81

Powell M., 1994, in Gomez S., Hennart J.-P., eds, Mathematics and Its Applications, Vol. 275, Advances in Optimization and Numerical Analysis. Springer, Netherlands, p. 51

Prevot M. L., Lequeux J., Prevot L., Maurice E., Rocca-Volmerange B., 1984, A&A, 132, 389

Repp A., Ebeling H., Richard J., 2016, MNRAS, 457, 1399

Salmon B. et al., 2015, ApJ, 799, 183

Sánchez C. et al., 2014, MNRAS, 445, 1482

Sawicki M., 2012, PASP, 124, 1208

Scoville N. et al., 2007, ApJS, 172, 1

Seaton M. J., 1979, MNRAS, 187, 73p

Speagle J. S., Eisenstein D. J., 2015, preprint (arXiv:1510.08073)

Wales D. J., Doye J. P. K., 1997, J. Phys. Chem. A, 101, 5111

This paper has been typeset from a TEX/LATEX file prepared by the author.